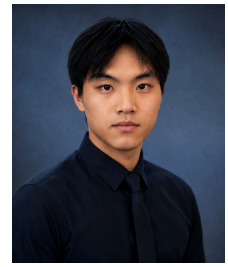


Tianlang Chen (Jackie), *B.S.E. Candidate*

✉ ctlang@umich.edu

🌐 <https://github.com/jackiectl>



Education

- 2025 – present 📖 **University of Michigan, Ann Arbor** Bachelor's Degree, **Data Science**
Relevant Coursework: *Machine Learning, Computer Graphics & Generative Models, Computational Linguistics, Algorithm & Data Structure, Probability & Statistics, Statistics & Data Analysis*
- 2023 – 2025/05 📖 **Shanghai Jiaotong University** Bachelor's Degree, **Mechanical Engineering**
Relevant Coursework: *Applied Linear Algebra, Dynamics System, Design & Manufacturing, Dynamics & Vibrations*

Internship

- 2024 – 2025 📖 **Mars Asia, Global IT Service Center** *WeCom full-stack development*
- Contributed to WeCom (Enterprise Wechat) full-stack secondary development using JavaScript, Node.js, and Vue.js to support internal enterprise communication and workflow tools.
 - Supported API implementation, interface development, data migration, and information retrieval features across internal service platforms.
 - Built and assisted with automation workflows to improve operational continuity during system migration.

Research

- 2026 – present 📖 **Large Language Model for Incentive Design in Mobility Systems** *University of Michigan, Ann Arbor*
- Investigating how large language models can support incentive mechanism design for shared mobility systems, including ridesharing, transit incentives, and vehicle-sharing.
 - Building LLM-guided workflows to generate and evaluate practical, interpretable incentive strategies for improving mobility system efficiency.
 - Studying the integration of learning-based methods and optimization techniques for data-driven transportation incentive design.

Research (continued)

2026 – present

■ **EPCOT Foundation Model: Scalable Data Preparation and Training Pipeline for TSS-to-Expression Prediction** *University of Michigan, Ann Arbor*

- **(1) Pseudobulk Data Processing:** Built an end-to-end pipeline transforming raw HPAP snMultiome data into pseudobulk format (per-donor, per-cell-type); converted ATAC single-cell 1 bp resolution sparse matrices into 600 kb-windowed HDF5 arrays, and RNA counts into CP10K-normalized 19,264-gene pseudobulk vectors with precise 7-cell-type stratification.
- **(2) Training Infrastructure Optimization:** Implemented 5 bit-equivalent optimizations (I/O prefetch with background threading, non-blocking pinned H2D transfers, gradient-accumulation without per-batch sync, memory-stall elimination, vectorized masking) achieving 2–5× per-epoch speedup while maintaining numerical equivalence to the baseline.
- **(3) Improved Generalization via Reset-Head Fine-Tuning:** Addressed backbone-reuse by loading pre-trained EPCOT weights and re-initializing the expression head from scratch, followed by staged backbone freezing—stage 1 (frozen backbone, fresh head learning), stage 2 (unfrozen backbone, low-LR adaptation) to prevent overfitting on limited new data.
- **(4) Discovered and Corrected TSS-Bin Offset Bug for Interpretability:** Identified and fixed a 50 kb positional shift in TSS-bin extraction (root cause: EPCOT's output retains full 600-bin length while the TSS predictor was indexing at bin 250 instead of 300), improving model interpretability and feature attribution accuracy.
- **(5) atac recover snp.**

2026 – present

■ **AI-Guided Environmental Sensing and Adaptive Monitoring** *University of Michigan, Ann Arbor*

- Investigating AI-guided sensing methods for sparse, dynamic Great Lakes environmental systems under limited sensor coverage, seasonal variability, and high uncertainty.
- Develop reproducible Software / AI workflows for environmental data ingestion, preprocessing, baseline reconstruction, uncertainty quantification, and visualization.
- Exploring sensor placement and adaptive sampling strategies that compare random, grid-based, and uncertainty-aware decision rules for environmental monitoring.
- Connect data-driven models with sensing-system design by generating heatmaps, uncertainty maps, placement recommendations, and performance curves for downstream hardware and deployment planning.

2026 – present

■ **Applets for Data Analytics in Materials Science** *University of Michigan, Ann Arbor*

- Developing web applets with JavaScript and a Python back end to support data analytics, data curation, and data sharing for the materials research community.
- Implementing interactive modules that enable users to upload datasets and analyze them using state-of-the-art models, theories, and data fitting methods.
- Collaborating with an existing programming team to build reusable analytics modules and workflow templates aligned with FAIR data principles.

Research (continued)

- 2026 – present
- Functional Genomic Representation Learning for Microbial Communities

University of Michigan, Ann Arbor

- Architecting a composable genomic representation framework that converts microbial genomes into interpretable, parameter-free functional encodings via proteogenic k-mer tokenization, enabling flexible integration with both classical statistics and large pre-trained sequence models (Evo, ESM, ProtT5).
 - Building abundance-weighted community embeddings from per-genome k-mer spectra, demonstrating a generic, model-agnostic approach to aggregating gene content into fixed-length community phenotypes; bridging mechanistic interpretability with foundation model representations.
 - Implementing leakage-safe feature engineering, comprehensive statistical validation (permutation tests, honest cross-validation, sample influence diagnostics), and compositional data transformations; architecting infrastructure for systematic ablation and reproducible science on high-dimensional, low-sample genomic data.

Skills

- Coding & Programming
- Python, C, C++, JavaScript, SQL, R, Matlab, Java, Bash, Git
- ML & Data
- PyTorch, scikit-learn, Gaussian Processes, Foundation-Model Fine-Tuning, HDF5
- LLM & Agentic AI
- Prompt Engineering, Structured Prompting, Retrieval-Augmented Generation, LLM Workflow Design, Agent-Guided Code Generation & Troubleshooting, Output Validation, Hallucination Mitigation
- Databases
- MySQL, SQLite.
- Web Dev
- HTML, CSS, JavaScript, Node.js, Vue.js.
- Languages
- English, Mandarin, Cantonese
- Tools
- Solidworks, Blender, MATLAB, Slurm / HPC

Additional Experience

Contests

- 2025/01
- The Mathematical Contest in Modeling (MCM-COMAP), Honorable Mention.

Teaching Assistant

- 2025/03
- TA for English Academic Writing
- 2025/05
- TA for General Physics